DOCUMENT RESUME

ED 443 873 TM 031 518

AUTHOR Cizek, Gregory J.; Trent, E. Roger; Crandell, Jan; Hirsch, Thomas; Keene, John

TITLE Research To Inform Policy: An Investigation of Pupil Proficiency Testing Requirements and State Education Reform Initiatives.

PUB DATE 2000-04-00

NOTE 29p.; Paper presented at the Annual Meeting of the American Educational Research Association (New Orleans, LA, April 24-28, 2000). Contains some areas of broken text.

PUB TYPE Reports - Evaluative (142) -- Speeches/Meeting Papers (150)

EDRS PRICE MF01/PC02 Plus Postage.

DESCRIPTORS Educational Change; *Educational Policy; *Elementary School Students; Grade 4; Grade 5; Intermediate Grades; *Reading Achievement; Reading Tests; *Student Promotion; *Test Use

IDENTIFIERS *Ohio Fourth Grade Proficiency Test; Reform Efforts

ABSTRACT

To investigate the potential effects of the "Fourth Grade Reading Guarantee," an Ohio policy that mandates that students will not be promoted to grade 5 unless they reach a given score on the Ohio Proficiency Test (OPT) in reading at grade 4, a series of studies was performed in the 1998-1999 and 1999-2000 school years. The extent of agreement among educators' judgments regarding students' academic preparation for work in the fifth grade, the relationships between these judgments and OPT scores, and the relationships between educators' judgments and policies intended to enhance educational quality were studied. Surveys in spring 1999 completed by teachers and principals yielded information on 6,065 (useable records), and 5,611 student records were obtained for the following school year, early fifth grade for these students. Teachers and principals tended to demonstrate a high degree of agreement about whether students were sufficiently prepared to enter fourth grade, but the operational definitions of adequate preparation differed in different school districts. Districts in which students generally perform less well had lower conceptualizations of competence. Educators' judgments were generally in line with students' actual pass/fail rates on the OPT. A fairly substantial discrepancy existed between the proportion of students that educators classified as not reading well enough to be academically successful in fifth grade and the proportion of students retained in grade, with the percentage of students judged underprepared more than 15 times greater than the percentage of students actually retained. Some areas of further research are outlined. (SLD)

Reproductions supplied by EDRS are the best that can be made
from the original document.

ED 443 873

Research to Inform Policy:

An Investigation of Pupil Proficiency Testing Requirements and State Education Reform Initiatives

G. Cizek

1

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)
This document has been reproduced as received from the person or organization originating it.
Minor changes have been made to improve reproduction quality.
Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

Gregory J. Cizek
University of North Carolina at Chapel Hill

E. Roger Trent
Ohio Department of Education

Jan Crandell
Ohio Department of Education

Thomas Hirsch
Assessment and Evaluation Services

John Keene
Assessment and Evaluation Services

TM031518

Paper presented at the annual meeting of the American Educational Research Association, April 2000, New Orleans, LA

BEST COPY AVAILABLE

Research to Inform Policy: An Investigation of

Pupil Proficiency Testing Requirements and State Education Reform Initiatives

Educational policies are often crafted in response to real or perceived crises. Unfortunately, well-supported justifications for policy creations or educational innovations do not always accompany proposals for reforms, nor does evidence demonstrating the likelihood of success of the proposals, nor is evidence regarding the effectiveness of the proposals disseminated after the intervention has run its course (Cizek & Ramaswamy, 1999). Too typically, alternative proposals for new educational innovations are introduced and implemented, obviating the need or impetus for studying the previous one, and the relationship between policy making, allocation of resources, educational innovations, and effects on key outcomes such as student achievement or educator practices remain only dimly understood.

This paper presents an examination of the interplay of some of these factors. As states struggle with using how to best allocate educational resources, large-scale, high-stakes tests are increasingly called upon to provide accurate information for informing policy decisions. This study addresses questions such as:

1) can external tests developed to assess proficiency in specific subject areas also be used to help inform promotion/retention decisions at a local level;

2) to what extent are educators able to respond to the requirements of a policy that asks them to accurately judge student preparation for promotion; and

3) to what degree do principals and teachers share common conceptions of adequate preparation?

Background and Objectives

This research examined the effects of an educational policy implemented in the state of Ohio, as a result of a series of legislative initiatives. The first relevant authority is Section 3301.0710 of the Ohio Revised Code (ORC), which authorizes the Ohio State Board of Education to prescribe tests, one in each of five subject areas (reading, writing, mathematics, science, and citizenship), to be administered for the purpose of measuring pupil achievement in the fourth, sixth, ninth, and twelfth grades. Hereafter, these tests are referred to as the Ohio Proficiency Tests (OPT). The legislature also authorized the State Board of Education to determine a score on each test that shall be deemed to demonstrate that students attaining such a score have achieved at least a specified level of proficiency in the measured skill, and required that students who score below the identified passing score at grade four be provided with appropriate intervention in grade five. Legislative initiatives such as these involving mandated pupil proficiency testing are increasingly common features of American educational policy making (Roeber, Bond, & Connealy, 1998).

As part of a recent initiative designed to stimulate education reform and enhance student achievement in elementary and secondary schools in the state, the 122nd Ohio General Assembly passed Amended Senate Bill 55, which was subsequently signed by the Governor on August 22, 1997. Among several changes, the legislation amended Section 3313.608 of the ORC in the following ways:

> (A) beginning with students who enter fourth grade in the school year that starts July 1, 2001, no city, exempted village, or local school district shall promote to fifth grade any student who fails to attain the score designated under division (a)(1) of Section 3301.0710 of the revised code on the test prescribed under that division to measure skill in reading, unless either of the following applies: (1) the pupil was excused from taking the test under division (c)(1) of section 3301.0711 of the revised code; (2) the pupil's principal and reading teacher agree that the pupil is academically prepared, as determined pursuant to the district policy adopted under section 3313.609 of the revised code, to be promoted to fifth grade.

These requirements are referred to colloquially as the "Fourth Grade Reading Guarantee." The requirements will apply to approximately 125,000 pupils who will take the 4th grade reading proficiency test (OPT-R4) in March, 2002.

It is relevant at this point to note that the current score required to pass the OPT-R4 was not validated to correspond with a minimum level of academic preparation for fifth grade. Rather, the focus of establishing the current passing score was on avoiding a specific type of incorrect decision; namely, avoiding incorrectly classifying students as *not* needing additional assistance. The approach used implicitly considered failing to correctly identify those students who might need additional services as a more serious "error" than failing to correctly identify those students who do not need such services. Thus, the current passing score was established, in part, to promote the likelihood that those pupils who might need remediation or special intervention services would receive that assistance. On the other hand, the Fourth Grade Reading Guarantee rests on the premise that the passing score used on OPT-R4 should represent a level of performance consistent with reading ability necessary for functioning at the fifth grade level. This approach implicitly places a different value on the different kinds of "errors." In other words, because the consequences of failing the OPT-R4 are now potentially much more serious than previously (i.e., retention vs. provision of additional services), the focus changes to preventing the more serious error of incorrectly identifying a student as not prepared for fifth grade.

In order to investigate potential effects of the Fourth Grade Reading G e, a series studies was conducted in the 1998-1999 and 1999-2000 school years. The specific research questions driving these studies addressed: 1) the extent of agreement among educators'(i.e., teachers' and principals') judgments regarding students' academic preparation for success work in the fifth grade; 2) relationships between educators' judgments and students' performance on the OPT-R4; and 3) relationships between educators' judgments and policies intended to enhance educational quality.

Method and Sample

Using stratified random sampling, a sample of classrooms from across the state was identified.

BEST COPY AVAILABLE

Stratification was performed to allocate districts across the state (n=611) to quintiles based on the districts' percent of fourth-grade students passing all five of the mandated 4th grade proficiency tests (reading, mathematics, science, citizenship, writing) administered in the 1997-1998 school year. Data were collected for all students in a selected classroom.

A first pair of surveys was mailed to teachers and principals in the spring of 1999. For each student in the selected classrooms, fourth grade teachers were asked to respond to the following question: "Does the student read well enough to be academically successful in the 5th grade?" Principals from the same schools were asked the identical question. Because the only alternative route for 4th grade students to be promoted to 5th grade under the Fourth Grade Reading Guarantee is concurring recommendations of the student's teacher and principal, investigation of these judgments was deemed particularly important. The spring surveys yielded information on 6576 students from 106 schools. These students were then matched with the results from the Spring 1999 OPT-R4 testing and only students for whom reading scaled scores were recorded were included; a 93% successful match rate was obtained. A final sample of 6065 student records was available for subsequent analyses.

In the fall of 1999, at the end of one complete grading period, the fifth-grade teachers at the same schools surveyed in the spring were asked the same question regarding the same students. Principals were asked to verify which students were still enrolled at their schools, and which students had not been promoted from 4th grade. Class rosters submitted to the 5th grade teachers permitted them to respond to the question regarding the students' reading proficiency, or to indicate that the student had been enrolled for less than the full grading period, that the student was no longer enrolled, or that the student had been retained. Responses were received for 5,843 students from 93 schools. The match with the grade 4 data produced 5,611 student records for a 96% match rate.

Results

Results of analyses conducted to address the three primary research questions are provided separately in the following subsections. Table 1 provides basic descriptive statistics showing the number

BEST COPY AVAILABLE

of respondents who answered "Yes" and "No" to the question "Does the student read well enough to be academically successful in the 5th grade?" The table also shows summary descriptive data on student performance in terms of scaled scores within each category.

Insert Table 1 about here.

Research Question 1: To what extent do educators concur in their judgments of students' preparation?

The first research question considered the extent to which teachers and principals would concur regarding 4th grade students' preparation in reading. The information in Table 1 provides some evidence bearing on this question. Note that the proportions of 4th graders judged to be adequately prepared by their 4th grade teachers and their principals were .806 and .808, respectively. A more precise answer to the question can be gained via crosstabulations, however. Table 2 shows the teachers and principals achieved approximately 94% overall agreement. Crosstabulations presented in Table 3 show the extent to which 4th and 5th grade teachers concurred in their judgments of students' preparation. These groups also showed fairly high agreement, with an overall consistency of 85%.

Finally, Table 4 shows the extent when concurring judgments of 4th grade teachers and principals were compared to the judgments of 5th grade teachers. In this case, too, agreement was fairly high, with 5th grade teachers agreeing overall in 82% of the cases in which the 4th grade teachers and principals provided concurring judgments.

Insert Tables 2-4 about here.

BEST COPY AVAILABLE

Research Question 2: What Are the Relationships Between Educators' Judgments and Students' Performance on the OPT-R4?

A comparison of educators' judgments and student performance (pass/fail) on the OPT-R4 is shown in Table 5. The table presents the crosstabulations of judgments and student performance in two ways: one in which students' pass/fail status on the test was determined using the passing scaled score originally required on the OPT-R4 (C_x = 200), and one using an increased passing standard adopted as part of efforts in Ohio to implement higher educational standards (C_x = 217). The table reveals moderately high levels of agreement between educators' judgments and students' performance in both cases, although the lower passing score produced greater agreement (i.e., the passing score of 200 produced proportion of agreement of 0.82; the passing score of 217 produced a proportion of agreement value of 0.71).

Insert Table 5 about here.

A different perspective on the relationship between educators' judgments and student performance can be seen in a comparison of the scaled scores associated with students judged to be sufficiently prepared for successful work in the 5th grade when these results are broken down by the strata over which sampling occurred. Recall that the sampling plan was conducted to form quintiles which effectively stratified school districts' performances across the state by overall student proficiency. The quintiles were combined to produce three levels: Quintile 1 = Level 1 (lowest overall performance), Quintiles 2-4 (middle performance level) = Level 2, and Quintile 5 = Level 3 (highest overall performance). Tables 6.1 to 6.3 contain the frequency distributions of the OPT-R4 scaled scores actually obtained by the 4th grade students whom their teachers had rated as adequately prepared for successful

performance in 5th grade. The dashed lines in the tables are placed at approximately the 50th percentile of each distribution and reveal that the scaled score associated with the median of these distributions increases as stratum increases.

Insert Table 6 about here.

A more precise illustration of this phenomenon can be seen in a plot of the frequency distributions of OPT-R4 scaled scores obtained by the 4th grade students whom their teachers had rated as adequately prepared for successful performance in 5th grade and the frequency distributions of the OPT-R4 scaled scores obtained by the students whom their teachers had rated as not adequately prepared. The graphs of these distributions illustrate how the intersections of the distributions can be used to show the differing conceptions of adequate performance for successful work in 5th grade vary across the strata. One such contrasting groups graph is found in Figure 1 for the Level 1 results. The data points are presented in the graph, as are curves fit to the points using a 6 degree polynomial. The intersection points for the smoothed distributions Level 1, Level 2, and Level 3 groups were 189, 196, and 203 respectively. These results suggest tl... "erent ope: 'onal definitions of adequate performance apply depending on the school district in which the student is enrolled. It appears that, in general, Level 1 students going into grade 5 are not likely to read as well as Level 3 students going into grade 5, and so on.

Insert Figure 1 about here.

BEST COPY AVAILABLE

Research Question 3: What is the Relationship between the Policy Initiative, Educator Judgments, and Student Performance?

In the Fall of 1999, fifth-grade teachers were asked the same questions as asked of the fourth grade teachers (i.e., "Does the student read well enough to be academically successful in 5th grade?). If the teacher responded "Yes," the response was coded as "1"; a "No" response was coded as "2"; if the teacher did not respond, a code of "9" was assigned. Because the principals surveyed in the fall were the same principals surveyed in the spring, they were not asked again to provide their judgments regarding the students' readiness. Instead, the principals were asked to indicate whether a student was promoted to fifth grade (coded "1"), retained in fourth grade (coded "2"), or no longer enrolled at the school (coded "4"); no response to this question was coded "9." The data indicating which students had been retained were used to examine the degree of correspondence between educators' judgments of student preparation, promotion/retention decisions, and implications of the policy initiative embodied in the Fourth Grade Reading Guarantee. Table 7 shows the results of this analysis. (Note: For students who were retained in the fourth grade or who no longer were enrolled in the school, the fifth-grade teachers gave no response which accounts for the large number of 9s in those ratings.)

Insert Table 7 about here.

Overall, principals reported that of the students for whom promotion/retention decisions were available, 5,029 of 4th grade students (98.8%) were promoted to 5th grade, while 62 (1.2%) were retained. These findings are particularly interesting in light of these same educators' judgments of student preparation for fifth grade reported earlier in this paper. Recall that, previously, principals and teachers had both judged that 19% of these 4th grade students did not read well enough to be academically successful in 5th grade (see Table 2). The current findings suggest that the

promotion/retention decision is being made based on additional information about students, or that other policies relevant to these decisions were in place in these school districts at the time of this study.

Conclusions and Recommendations for Future Research

The careful sampling plan designed for this study and the very high response rates from teachers and principals suggests that we can have a fairly high degree of confidence in the initial findings from this study. Naturally, these findings apply primarily to the educational system, practices, personnel, and policies in the state of Ohio. However, to the extent that similar policies are in effect or contemplated in other, similar states, this research may provide some guidance for researchers, educators, and policy makers. The following sections contain a summary of the results presented in this paper, a summary of questions left unanswered, and an outline of follow-up research on this topic planned for the 2000-2001 school year.

What We Know and What We Don't Know

Some of the conclusions that we believe are supported by this research include the following:

1) teachers and principals tend to demonstrate a high degree of agreement when judging whether students are sufficiently well-prepared in reading at the fourth grade for successful performance in the fifth grade;

2) educators' operational definitions of adequate preparation in reading for successful work in the fifth grade vary in relation to the school district in which a student is enrolled. Districts in which students, on average, perform less well overall on state measures of proficiency tend to have conceptualizations of competence that are lower than those held by educators in school districts in which students, on average, perform at higher levels;

3) educators' judgments regarding adequacy of preparation are generally in line with students' actual pass/fail status on the state reading proficiency test;

4) there exists a fairly substantial discrepancy between the proportion of students that educators

classify as *not* reading well enough to be academically successful in 5th grade and the proportion of students retained in grade, with the percentage of students judged to be under-prepared more than 15 times greater than the percentage of students actually retained. From an educational policy perspective, one effect of the Fourth Grade Reading Guarantee could be to change this, such that (potentially) a greater proportion of students would be retained in fourth grade as a result of failing to attain the required score on the OPT-R4.

Just as some questions have been answered (at least tentatively) by this research, the answers to other questions remain unclear, and other questions arise. Among the issues that are unresolved include:

1) the process by which teachers and principals obtained the fairly high degree of agreement witnessed in this study. Although it is plausible that some principals are sufficiently familiar with the reading skills of the 4th grade students enrolled in their schools, it is not likely that principals generally have this kind of familiarity. It seems reasonable to suggest that principals' judgments were generated, to a great degree, based on their inclinations to accept teachers' professional judgments. Indeed, some anecdotal evidence obtained during this research suggests that this is the case. If so, this suggests that one aspect of the Fourth Grade Reading Guarantee (i.e., the aspect permitting students who do not obtain the required score on the OPT-R4 to be promoted as long as the student's teacher and principal agree that the student is adequately prepared) may not be functionally incorporating two different sources of information;

2) the additional sources of information being used by educators to make their judgments of student preparation in reading. It would be of interest to learn what kinds of data--beyond students' OPT-R4 test scores--could be used to improve predictions of educators' judgments regarding students' preparation for success in 5th grade;

BEST COPY AVAILABLE

3) the reasons and consequences of differential conceptualizations of adequate preparation for fifth grade. We do not know why levels of reading proficiency judged to be adequate for successful work in 5th grade appear to be dependent upon the district in which a student attends school. Nor do we know whether these "expectations" actually affect students' achievement in the manner of a self-fulfilling prophecy (see Rosenthal & Jacobson). It is reasonable to wonder, however, to what extent observed differential performance on state-mandated student proficiency tests in the fourth--and subsequent grades, including tests that must be passed in order to obtain a high school diploma--may be due these differing oprationalizations of what is necessary for student success. Or, it may be "success" itself is defined differently across the different strata sampled in this study. In any case, these differences suggest concerns about equitable student educational experiences and warrant further investigation; and

4) what other factor(s) educators took into account when making promote/retain decisions. It is clear that promotion/retention decisions currently are not strongly related to students' performance on the state-mandated reading test taken in the fourth grade. It would be of interest to learn what additional sources of information or policy bear on this decision. And, given that some students *were* retained and others were promoted who were judged to be underprepared, it would be interest to learn what, if any, diagnostic and remedial strategies or programs were instituted to assist retained and underprepared-but-promoted students achieve greater levels of reading proficiency.

Follow-up and Extensions of the Initial Studies

It is anticipated that a series of studies will be conducted to further clarify the issues raised by this research. The series consists of three elements, the intent and suggested characteristics of which are described separately below.

BEST COPY AVAILABLE

Element 1: Replication

As mentioned previously, the initial studies suggested tentative conclusions about some of the complex relationships among educators' judgments and between those judgments and actual student performance. Accordingly, as a first step, it would seem prudent to verify the initial findings in a replication study in which: feedback from the initial surveys sent to principals and teachers is reviewed; the survey design, sampling plan, etc., are adjusted as necessary; and the original study is replicated using a different sample of teachers, principals, schools, to serve as a "cross validation" of initial results.

Element 2: Longitudinal Data Collection

In addition to a replication, the initial study can be extended longitudinally to examine the perspectives of the same fifth grade teachers who had been surveyed in the original sample *after* they have been exposed to their students for a longer period of time. Recall that, in the original sample, the fifth grade teachers were asked after one grading period to respond to the following question: "Does the student read well enough to be academically successful in the 5th grade?" Of interest is determining how the teachers' perceptions may change with additional experience with each child. The same fifth grade teachers could be questioned regarding the same students they originally judged, but with data collection taking place near the end of the current (i.e., 1999-2000) school year. The information from this data collection would shed additional light on the initial findings. For example, it may be found that fifth grade teachers who had originally judged some students to be unprepared for successful performance in the fifth grade would evaluate those students as performing successfully as the end of the year approaches (or, conversely, they may evaluate some students whom they had judged to be prepared to be underperforming). In either case, the degree of stability of the teachers' evaluations would provide a criterion to consider when evaluating the correspondence between the fourth- and fifth-grade teachers' judgments as observed in the original study. This element could include asking the fifth-grade teachers to describe the aspects of classroom instruction that fostered or prevented the students' success or failure. For example, for students that the teachers' judged to be performing successfully overall in the fifth

grade, it would be of interest to know:

1) the ways in which fifth-grade teachers operationalize "successful performance in the fifth grade" (e.g., turns in required assignments, works independently, etc.);

2) which students were functioning successfully overall without special intervention;

3) for students who required special assistance to function successfully, which interventions were being implemented; and

4) for students who were judged to *not* be functioning successfully, in which areas the teachers judged that the students needed special assistance and what kinds of assistance they deemed would be most effective.

A number of potentially highly informative analyses can be conducted using the data that would result from the longitudinal data collection. One particularly informative analysis would involve examination of the judgments obtained in step 4 (above) in light of the teachers' initial judgments regarding students' preparation. Such an analysis would shed light on intervention/instructional strategies that would enhance the prospects of success for student identified in the fourth grade as potentially at-risk.

Element 3: Case Studies

It is important for teachers, administrators, policy makers, and parents to understand how best to approach student success and failure on the state-mandated tests given the high stakes associated with them. We begin with the assumption that some students will be retained in fourth grade as a result of application of the policy embodied in the Fourth Grade Guarantee (i.e, as a result of being unable to attain the passing score on OPT-R4). Focussed case studies of purposefully selected students and teachers could be used to provide some data relevant to the following questions:

For students, parents and teachers of students retained in fourth grade:

1) What kinds of instruction/intervention were provided to students? Was instruction modified or different from what the students had received in the previous year? Were any additional interventions provided?

2) If additional services were provided, how can the logistics, costs, etc., of those services be expressed?

3) If special assistance was provided, what kinds of special assistance do teachers and/or parents perceive to have been most beneficial? Least beneficial? What kinds of additional assistance, if available, do teachers believe would be most efficacious in providing remediation for the retained students? What are the barriers to providing these additional resources?

4) What benefits do the retained students, their teachers and parents see as a result of the retention and/or the additional services (if applicable)?

5) In particular, in what ways has the student's proficiency in reading been affected by retention?

6) What have been some of the unforseen or unintended consequences of retention (either positive or negative), including but not necessarily limited to academic consequences?

7) How do the persons involved evaluate the cost/benefit of retention?

For teachers, students, and parents of students promoted to 5th grade:

As mentioned previously, the Fourth Grade Reading Guarantee permits students who failed to obtain the required score on the OPT-R4 to be promoted to 5th grade if "the pupil's principal and reading teacher agree that the pupil is academically prepared ... to be promoted to fifth grade." Thus, two types of students will have been promoted to fifth grade: those who obtained the required score on OPT-R4; and those who failed to obtain the required score but were promoted based on teacher and principal concurrence regarding preparation. Gathering information about both of these groups of students, their teachers, and their instructional programs will assist in determining which types of interventions are best

BEST COPY AVAILABLE

suited to promoting success in fifth grade for students entering that grade with deficiencies identified on the OPT-R4. It would likely be most informative if the case studies focussed on students from the two groups whose scores were just above and just below the passing score. Among the questions that would be relevant for this element are:

1) What kinds of instruction/intervention were provided to the students by their fifth grade teachers? Was instruction modified or different from what other fifth grade students received (i.e., was instruction individualized based on knowledge of the students OPT-R4 performance? In what ways? Who were any additional interventions provided?)

2) If additional services were provided, how can the logistics, costs, etc., of those services be expressed?

3) If special assistance or intervention programs have been implemented for these students, what kinds of special assistance do teachers and/or parents perceive to have been most beneficial? Least beneficial? What kinds of additional assistance, if available, do teachers believe would be most efficacious in providing remediation for the promoted students? What are the barriers to providing these additional resources?

4) What benefits do the promoted students, their teachers and parents see as a result of the promotion as compared to retention?

5) How has the student's proficiency in reading developed? Is the student functioning successfully in other academic areas?

6) What have been some of the unforseen or unintended consequences of promotion (either positive or negative), including but not necessarily limited to academic consequences?

7) How do the persons involved evaluate the cost/benefit of promotion?

One additional group may be of interest for the case study approach. That group consists of students who scored above the passing mark on OPT-R4, but who were identified by their teachers as not

sufficiently prepared for fifth grade. Although answers to the above questions for this group would be of great interest, available data suggest that there are so few of these students that it may be impractical or inconclusive even if they were to be included in the follow-up study.

References

Cizek, G. J., Ramaswamy, V. (1999). American educational policy: Constructing crises and crafting solutions. In G. J. Cizek (Ed.), *Handbook of educational policy* (pp. 498-519). San Diego, CA: Academic.

Roeber, E., Bond, L., & Connealy, S. (1998). *Annual survey of state student assessment programs.* Washington, DC: Council of Chief State School Officers.

Rosenthal, R., & Jacobson, L. (1968). *Pygmalion in the classroom.* New York: Holt, Rinehart & Winston.

Table 1

Summary Statistics for Grade 4 Teachers, Principals, and Grade 5 Teachers and Student Scaled Scores

	Grade 4 Teachers		Principals		Grade 5 Teachers	
	Yes	No	Yes	No	Yes	No
N of cases	4750	1146	4844	1154	4088	801
Proportion	0.806	0.194	0.808	0.192	0.836	0.164
Minimum	134.000	147.000	159.000	147.000	134.000	147.000
Maximum	271.000	258.000	271.000	258.000	271.000	258.000
Range	137.000	111.000	112.000	111.000	137.000	111.000
Mean	222.681	202.810	222.388	203.379	223.773	202.830
Variance	266.402	233.417	268.037	252.522	248.096	236.021
Std. Deviation	16.322	15.278	16.372	15.891	15.751	15.363
Std. Error	0.237	0.451	0.235	0.468	0.246	0.543
Skewness	-0.012	-0.178	0.037	-0.057	0.077	-0.109
Kurtosis	0.701	0.156	0.477	0.198	0.805	0.186

Table 2

Crosstabulations of Grade 4 Teachers' and Principals' Judgments

Grade 4 Teachers' Judgments	Principals' Judgments Consistency = 0.94		
	Yes	**No**	**Total**
Yes	4586	172	4758
	.78	.03	.81
No	156	970	1126
	.03	.16	.19
Total	4742	1142	5884
	.81	.19	1.00

Table 3

Crosstabulations of Grade 4 and Grade 5 Teachers' Judgments

Grade 4 Teachers' Judgments	Grade 5 Teachers' Judgments Consistency = 0.85		
	Yes	**No**	**Total**
Yes	3707	333	4040
	.76	.07	.83
No	366	455	821
	.08	.09	.17
Total	4073	788	4861
	.84	.16	1.00

Table 4

Crosstabulations of Grade 4 Teachers' and Principals' Judgments by Grade 5 Teachers' Judgments

Judgments for Grade 4 Teachers and Principals	Grade 5 Teachers' Judgments Consistency = 0.82		
	Yes	**No**	**Total**
Teacher =Yes Principal = Yes	3609	283	3892
Proportion	.74	.06	.80
Teacher =No Principal = No	295	396	691
Proportion	.06	.08	.14
Teacher =Yes Principal = No	97	50	147
Proportion	.02	.01	.03
Teacher =No Principal = Yes	70	59	129
Proportion	.01	.01	.03
Total	4071	788	4859
Proportion	.84	.16	1.00

Table 5

Crosstabulation of 4th Grade Teachers' Judgments and the Student Performance

Passing score	Grade 4 Teachers' Judgments		
	Yes	No	Total
200	(overall agreement = 0.82)		
Pass	4379	671	5050
	.74	.11	.86
Fail	371	475	846
	.06	.08	.14
217	(overall agreement = 0.71)		
Pass	3242	227	3469
	.55	.04	.59
Fail	1508	919	2427
	.26	.16	.41

Table 6

Frequency Distributions for 4th Grade Teachers Classifying Students as "Prepared" by Level

Table 6.1- Level 1

Count	Cumulative Count	Percent	Cumulative Percent	Scaled Score
4	4	.4	.4	159.0
1	5	.1	.5	163.0
2	7	.2	.6	165.0
2	9	.2	.8	168.0
3	12	.3	1.1	170.0
1	13	.1	1.2	173.0
3	16	.3	1.5	175.0
6	22	.6	2.0	177.0
5	27	.5	2.5	179.0
4	41	1.3	3.8	181.0
8	49	.7	4.5	183.0
2	61	1.1	5.6	185.0
2	73	1.1	6.7	186.0
7	80	.6	7.4	188.0
6	96	1.5	8.9	190.0
7	123	2.5	11.4	192.0
5	138	1.4	12.8	194.0
9	157	1.8	14.5	196.0
8	185	2.6	17.1	197.0
8	213	2.6	19.7	199.0
7	250	3.4	23.1	201.0
7	277	2.5	25.6	203.0
2	319	3.9	29.5	205.0
4	363	4.1	33.5	207.0
6	419	5.2	38.7	210.0
4	473	5.0	43.7	212.0
7	540	6.2	49.9	214.0
---	---	---	---	---
3	623	7.7	57.6	217.0
2	705	7.6	65.2	220.0
1	786	7.5	72.6	224.0
5	871	7.9	80.5	228.0
4	955	7.8	88.3	232.0
3	1018	5.8	94.1	238.0
1	1059	3.8	97.9	246.0
2	1081	2.0	99.9	258.0
1	1082	.1	100.0	271.0

Table 6.2 - Level 2

Count	Cumulative Count	Percent	Cumulative Percent	Scaled Score
2	2	.1	.1	168.0
1	3	.0	.1	175.0
2	5	.1	.2	177.0
3	8	.1	.4	179.0
3	11	.1	.5	183.0
7	18	.3	.8	185.0
5	23	.2	1.1	186.0
7	30	.3	1.4	188.0
5	35	.2	1.6	190.0
10	45	.5	2.1	192.0
13	58	.6	2.7	194.0
22	80	1.0	3.8	196.0
29	109	1.4	5.1	197.0
20	129	.9	6.1	199.0
30	159	1.4	7.5	201.0
44	203	2.1	9.5	203.0
58	261	2.7	12.3	205.0
75	336	3.5	15.8	207.0
91	427	4.3	20.1	210.0
109	536	5.1	25.2	212.0
161	697	7.6	32.7	214.0
168	865	7.9	40.6	217.0
220	1085	10.3	51.0	220.0
220	1305	10.3	61.3	224.0
257	1562	12.1	73.4	228.0
204	1766	9.6	82.9	232.0
168	1934	7.9	90.8	238.0
119	2053	5.6	96.4	246.0
61	2114	2.9	99.3	258.0
15	2129	.7	100.0	271.0

Table 6.3 - Level 3

Count	Cumulative Count	Percent	Cumulative Percent	Scaled Score
1	1	.1	.1	175.0
1	2	.1	.2	192.0
2	4	.2	.4	194.0
7	11	.7	1.1	196.0
3	14	.3	1.4	197.0
3	17	.3	1.7	199.0
8	25	.8	2.5	201.0
8	33	.8	3.3	203.0
16	49	1.6	4.9	205.0
13	62	1.3	6.2	207.0
10	72	1.0	7.2	210.0
29	101	2.9	10.1	212.0
39	140	3.9	14.0	214.0
53	193	5.3	19.3	217.0
90	283	9.0	28.3	220.0
112	395	11.2	39.5	224.0
126	521	12.6	52.1	228.0
149	670	14.9	67.0	232.0
137	807	13.7	80.7	238.0
108	915	10.8	91.5	246.0
68	983	6.8	98.3	258.0
17	1000	1.7	100.0	271.0

Table 7

Relationships between Fifth Grade Teachers' Judgments of Student Preparation and Promotion/Retention Decisions

Code	Grade 5 Teachers	Principals
1	4088	5029
2	801	62
4	---	491
9	722	29
Total	5611	5611

Figure 1

Contrasting Groups Plot of Level 1 "Yes" and "No" Frequencies